봉수 실증분석 논문 · 봉수 자체 지수 시리즈 ① · v1.5 기준

봉수 AI 개입지수(BAII) 실증분석 — 판매 상품 72 종의 AI 개입 성분 표시와 가중치 민감도

"얼마나 AI 가 만들었고, 얼마나 사람이 확인했나"를 숫자로 먼저 밝히는 성분 표시의 첫 실측



핵심 주제

AI 개입지수 · 성분 표시 · 출처원장 · 분산 분해 · 디리클레 민감도 · 개선 경로 · 표준가격 매트릭스

저자: 조봉수(봉수 SSGlobal(주) 대표이사, 필명). 투자 권유 아님.

정식본 2026 년 10 월

목차

국문 요약 **3**
1. 서론 **4**
1.1 문제 제기 4
1.2 연구 질문 4
2. 지수 정의와 자료 **4**
2.1 산식과 등급 4
2.2 자료 5
3. 결과 **5**
3.1 분포 — 두 무리 5
3.2 무엇이 점수를 만드나 6
3.3 가중치 민감도 6
3.4 개선 경로 — 무엇을 고치면 좋아지나 7
4. 논의 **8**
4.1 지수가 보여 준 봉수의 현재 8
4.2 정책 제안 8
5. 한계 **8**
6. 결론 **8**
참고문헌 **8**
부록 **A.** 인공지능 이용 고지 **9**
부록 **B.** 재현 방법 **9**

국문 요약

봉수는 AI 기반 정보플랫폼 회사로서 상품마다 **AI** 가 얼마나 개입했는지 를 먼저 밝힌다. 봉수 AI 개입지수(Bongsu AI Involvement Index, BAII)는 상품을 판정 단위(문장·문항·도표)로 나누어 다섯 성분을 0~1 로 재고, 가중 합산해 0~100 점으로 나타낸다.

- c1 무출처: 가중치 30
- c2 추정·예측: 가중치 25
- c3 AI 생성: 가중치 20
- c4 재현 불가: 가중치 15
- c5 수치 미검수: 가중치 10

본 연구는 2026 년 10 월 5 일 bongsu.kr/jisu 에 공개한 판매 상품 **72** 종 전체 의 실제 산출표로 이 지수를 실증분석하였다.

주요 결과는 다음과 같다.

① 두 무리로 갈린다. 상품은 두 무리로 뚜렷하게 나뉘었다.

무리	상품 수	**BAII** 평균	등급
해설·논문 상품	50 종	46.6	B·C
예측 위주 예상문제	22 종	99.1	F

두 무리의 차이는 맨-휘트니 검정 $p<0.001$ 이다.

② 모든 상품이 똑같이 **30** 점을 받는다. 72 종 전부가 AI 생성 100%(c3)이고, 사람이 원자료와 대조한 수치가 0%(c5)이다. 이 두 성분이 모든 상품에 똑같이 30 점을 더한다. 그래서 상품 간 차이**(**분산**)**는 **c1·c2·c4** 세 성분이 **100%** 만든다. 성분별 분산 기여는 c2 47%, c4 29%, c1 25%이다.

③ 순위는 가중치에 흔들리지 않는다. 가중치를 디리클레 분포로 5,000 번 흔들어도 상품 순위의 스피어만 상관은 중앙값 0.999(하위 5% 0.993)였다. 순위는 가중치 선택에 거의 흔들리지 않는다.

④ 등급은 흔들린다. 해설 상품의 점수가 B/C 경계(45 점) 바로 근처에 몰려 있어, 가중치를 흔들면 상품의 31%(중앙값)가 다른 등급으로 옮겨 갔다. 등급 경계 근처의 판정은 가중치에 민감하다.

⑤ 개선 경로는 수치 대조다. 사람이 수치를 원자료와 대조하면(c5=0) 41 종이 C 에서 B 로 오르고 평균 표준가가 1,933 원에서 2,389 원으로 오른다. 원자료·코드를 공개하는 것(c4=0)은 이미 공개 중인 해설 상품에는 변화가 없었다.

결론적으로 BAII 는 순위 지표로는 견고하고, 등급 경계에서는 민감하다. 가장 값싼 개선 경로는 "사람이 수치를 원자료와 대조하는 것"이다. 지금 봉수 상품의 '사람 검증 사실' 비율이 0%라는 사실도 그대로 공개한다.

주제어: AI 개입지수, 성분 표시, 출처원장, 가중치 민감도, 정보상품 가격

1. 서론

1.1 문제 제기

생성형 AI 로 만든 정보상품이 빠르게 늘고 있지만, 소비자는 상품의 어느 부분이 사실 확인을 거쳤고 어느 부분이 AI 의 추정·생성물인지 알 수 없다. 출판계에서는 AI 이용 사실을 묻고 나서야 인정하는 일이 생겼다. 봉수는 반대로 묻기 전에 밝히는 쪽 을 택했고, 그 밝힘을 말이 아니라 숫자로 표시하려고 BAII 를 만들었다.[1]

1.2 연구 질문

- 연구 질문 1: 실제 판매 상품의 BAII 는 어떻게 분포하는가?
- 연구 질문 2: 상품 간 차이는 어느 성분이 만드는가?
- 연구 질문 3: 저자가 정한 가중치를 바꾸면 순위와 등급은 얼마나 흔들리는가?
- 연구 질문 4: 무엇을 고치면 등급과 가격이 가장 크게 좋아지는가?

2. 지수 정의와 자료

2.1 산식과 등급

표 1. BAII 성분과 가중치

성분	계산	가중치
c1 무출처	1 − 출처 있는 단위 ÷ 전체 단위	30
c2 추정·예측	추정·예측 단위 ÷ 전체 단위	25
c3 AI 생성	AI 가 생성한 단위 ÷ 전체 단위	20
c4 재현 불가	1 − (원자료 공개 + 코드 공개) ÷ 2	15
c5 수치 미검수	1 − 사람이 대조한 수치 ÷ 전체 수치	10

BAII = 30·c1 + 25·c2 + 20·c3 + 15·c4 + 10·c5 이며, 등급은 S(15 이하)·A(30)·B(45)·C(60)·D(80)·F(80 초과)이다. 등급은 품질 평가가 아니라 성분 표시 이다. 예측문제처럼 처음부터 미래를 다루는 상품은 F 가 정상이다. 표준가격 매트릭스는 등급에 따라 S 5,000 · A 4,000 · B 3,000 · C 2,200 · D 1,500 · F 1,000 원이다.

[1] 산식·등급표는 「봉수 AI 개입지수와 봉수 예측지수를 이용한 AI 정보상품의 표시·가격 표준 연구」(2026-10-05)와 공지 v1.5 에 처음 공개되었다. 이 논문은 그 지수를 실제 판매 상품 72 종에 적용한 첫 실증분석이다.

2.2 자료

자료는 2026 년 10 월 5 일 공개한 「봉수지수_공개표_20261005.csv」 72 행 전체다(bongsu.kr/jisu 에서 누구나 내려받을 수 있다). 상품은 고전 논술 논문 40 종, 논술·수능·안보 논문 10 종, 예상문제 22 종이며, 판정 단위는 상품당 13~1,057 개(중앙값 405.5)다. 공개표의 BAII 를 성분과 가중치로 다시 계산하면 최대 차이 0.06 점(반올림 오차)으로 일치하였다.[2]

3. 결과

3.1 분포 — 두 무리

표 2. 유형별 BAII

구분	상품 수	평균	최소	최대	무출처(%)	추정예측(%)	재현불가(%)
고전 논술	40	46.8	44.0	52.8	55.0	1.0	0.0
논문	10	46.0	42.8	48.9	53.0	0.5	0.0
예상문제	22	99.1	97.7	100.0	96.9	100.0	100.0

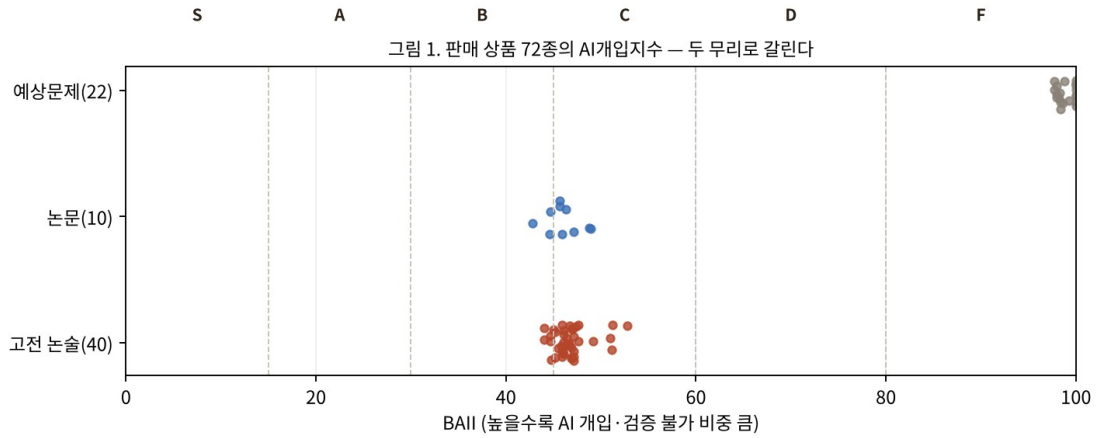


그림 1. 판매 상품 72종의 AI개입지수 — 두 무리로 갈린다

해설·논문 상품 50 종은 42.8~52.8 점에 몰려 B(9 종)·C(41 종)이고, 예상문제 22 종은 97.7~100 점으로 전부 F 다. 판정 단위 수와 BAII 의 순위상관은 −0.58(p<0.001)이었는데, 이는 규모 자체의 효과라기보다 예상문제가 짧은(평균 28 단위) 상품이기 때문이다. 중간 등급**(D)**과 상위 등급**(S·A)**은 하나도 없다.

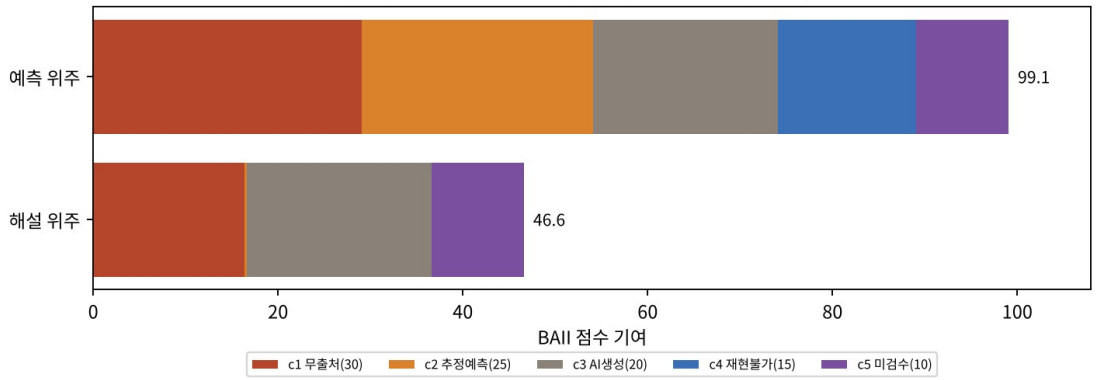
[2] 재계산 코드는 부록 B 의 fd_three.py 1 절. 공개표의 성분값이 소수 첫째 자리로 반올림되어 있어 생기는 차이다.

3.2 무엇이 점수를 만드나

표 3. 성분별 기여(분산 기여는 공분산 포함 분해, 합 100%)

성분	평균 점수 기여 비중	상품 간 분산 기여
c1 무출처	32.4%	24.5%
c2 추정·예측	12.4%	47.0%
c3 AI 생성	31.9%	0%
c4 재현 불가	7.3%	28.5%
c5 수치 미검수	16.0%	0%

그림 2. 무엇이 점수를 만드나 — AI생성·미검수 30점은 모든 상품에 똑같이 붙는다



c3(AI 생성)과 **c5(**수치 미검수**)**는 **72** 종 모두 **100%** 다. 봉수 상품은 전부 AI(Claude)가 생성했고, 수치를 사람이 원자료와 하나하나 대조한 기록이 아직 없다. 이 두 성분은 모든 상품에 똑같이 30 점을 더하므로 순위에는 영향이 없지만 절대 수준을 **30** 점 끌어올린다**.** 그래서 해설 상품도 S·A 가 될 수 없다. 상품 간 차이는 추정·예측 비중(c2), 원자료·코드 공개(c4), 출처(c1)가 만든다.

3.3 가중치 민감도

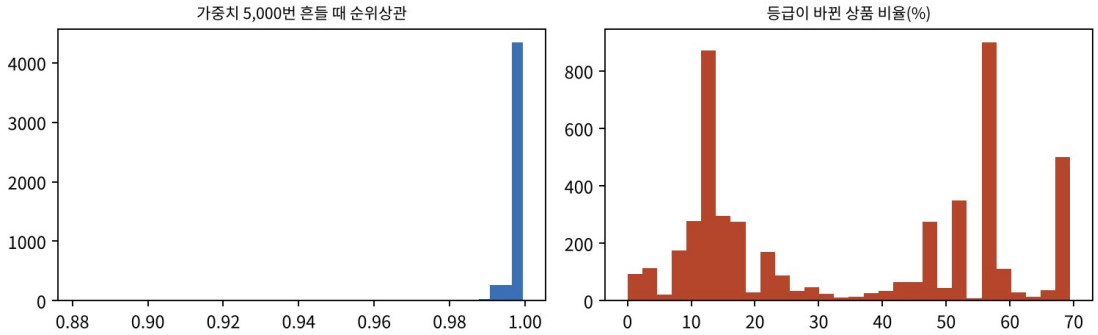
가중치(30·25·20·15·10)는 저자의 판단으로 정한 값이다. 이 선택이 결과를 좌우하는지 보려고 현재 가중치를 중심으로 하는 디리클레 분포(집중도 20)에서 가중치 5,000 세트를 뽑아 다시 계산하였다.[3]

표 4. 가중치를 5,000 번 흔든 결과

지표	값
순위상관(스피어만) 중앙값	0.999
순위상관 하위 5%	0.993
등급이 바뀐 상품 비율 중앙값	31.2%
등급이 바뀐 상품 비율 95 분위	69.4%

[3] 디리클레 분포는 합이 일정한(여기서는 100) 가중치 묶음을 무작위로 뽑는 표준 방법이다. 집중도 20 은 각 가중치가 대략 ±40% 범위에서 움직이게 한다. 씨앗 20261005 로 고정했다.

그림 3. 가중치 민감도(디리클레, 중심=현재 가중치)



순위는 거의 그대로였다. 무엇이 더 AI 개입이 큰 상품인지에 대한 판단은 가중치에 흔들리지 않는다. 반면 등급은 꽤 흔들렸다. 해설 상품 50 종이 B/C 경계(45 점)의 −2.2~+7.8 점 범위(42.8~52.8 점)에 몰려 있기 때문이다(경계 ±2 점 이내만 34 종). 등급과 그에 연결된 가격**(B 3,000 원 vs C 2,200 원)**은 경계 근처에서 가중치 선택에 민감하다**.** 이 점은 표준가격 운영의 한계로 공개해야 한다.

3.4 개선 경로 — 무엇을 고치면 좋아지나

표 5. 반사실 시나리오(가정한 개선을 넣어 다시 계산 — 실제 개선이 아님)

시나리오	평균 **BAII**	등급 분포	평균 표준가	등급이 오른 상품
현재	62.6	B9 C41 F22	1,933 원	—
수치를 사람이 원자료와 대조 (c5=0)	52.6	B50 F22	2,389 원	41
원자료·코드 모두 공개(c4=0)	58.1	B9 C41 F22	1,933 원	0
출처 단위 20%p 늘림	56.6	B47 C3 F22	2,356 원	38
대조 + 공개 동시 (c4=c5=0)	48.1	B50 D22	2,542 원	63

가장 효과가 큰 것은 사람의 수치 대조 였다. 해설 상품 41 종이 C 에서 B 로 올라 평균 표준가가 456 원 오른다. 원자료·코드 공개는 해설 상품에서는 이미 되어 있어(c4=0) 변화가 없고, 예상문제는 대조와 공개를 함께 해야 F 에서 D 로 오른다. 즉 BAII 는 봉수에게 **"**사람 검수에 시간을 쓰면 정당하게 값을 더 받을 수 있다**"** 는 신호를 준다. 지수가 속임수가 아니라 개선 유인으로 작동하도록 설계된 셈이다.

4. 논의

4.1 지수가 보여 준 봉수의 현재

BAII 의 가장 중요한 발견은 숫자 자체보다 **"사람 검증 사실 0%"** 라는 사실이다. 봉수는 이것을 숨기지 않는다. 오히려 공개표에 그대로 올렸다. 고객은 이 표를 보고 AI 가 만든 해설을 그 값에 살지 스스로 판단한다. 봉수에게는 다음 할 일(수치 대조 기록)이 분명해진다.

4.2 정책 제안

- 경계 표시: 등급 경계 ±2 점 안의 상품은 "경계 등급"으로 함께 표시한다.
- 검수 칸 신설: 출처원장에 '사람 대조' 칸을 실제로 채우는 작업을 시작한다(사람 시간은 원가기록에 함께 남긴다).
- 가중치 재검토: 1 년 뒤 고객 선택 자료(어느 등급이 실제로 팔렸나)로 가중치를 다시 검토하고, 바꾸면 새 판으로 공개한다.

5. 한계

- 사례 범위: 72 종 모두 봉수 상품이며 대부분 논술 분야다(69 종). 다른 판매자·분야에 그대로 일반화할 수 없다.
- 판정 방식: 판정 단위의 출처·추정 여부는 AI 가 규칙에 따라 판정했다. 사람이 표본을 뽑아 다시 확인하는 일은 아직 하지 않았다.
- 가중치의 성격: 가중치는 실증으로 추정한 값이 아니라 저자가 정한 제안값이다. 본 연구는 그 선택의 민감도만 보였다.
- 반사실 시나리오: 개선 효과는 계산상의 결과이며 실제 개선이 일어난 것이 아니다.

6. 결론

BAII 는 판매 상품 72 종에서 해설 상품과 예측 상품을 뚜렷이 갈라냈다. 상품 순위는 가중치 선택에 거의 흔들리지 않았지만, 등급은 경계 근처에서 흔들렸다. 모든 상품이 AI 생성 100%, 사람 수치 대조 0%라는 현재 상태를 지수가 그대로 드러냈다. 가장 효과적인 개선은 사람의 수치 대조다. 봉수는 이 결과를 공개하고, 검수 기록을 쌓아 다음 산출에서 달라진 숫자로 다시 보여 줄 것이다.

참고문헌

- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.

- Saisana, M., Saltelli, A. and Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. *Journal of the Royal Statistical Society A*, 168(2), 307–323.
- OECD/JRC (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing.
- 조봉수 (2026). 봉수 AI 개입지수와 봉수 예측지수를 이용한 AI 정보상품의 표시·가격 표준 연구. 봉수 SSGlobal(주).
- 봉수 SSGlobal(주) (2026). 공지 — 봉수 표준가격 정책 v1.5 개정(예측 성과 보상). bongsu.kr.

부록 A. 인공지능 이용 고지

본 논문의 자료 처리·통계 계산·그림·초고 문장은 인공지능(Claude)과 함께 작성하였다. 지수의 원칙과 공개 결정은 저자가 했다. 상품 판정도 AI 가 규칙에 따라 했으며, 그 결과 '사람 검증 0%'를 그대로 적었다. 코드와 원자료를 함께 공개하므로 누구나 같은 숫자를 다시 만들 수 있다.

부록 B. 재현 방법

1. 01_로우데이터 에 봉수지수_공개표_20261005.csv 를 둔다(bongsu.kr/jisu 에서 받을 수 있음).
2. 02_파이썬 에서 python -X utf8 fd_three.py 를 실행한다.
3. outputs 에 BAII_유형별.csv, BAII_개선경로.csv, key_results_3 지수.json 이, 04_논문/fig 에 그림이 생긴다.
4. 이용 조건: 산식·등급 기준은 CC BY 4.0, 코드는 MIT 라이선스.